

# IDENTIFICATION AND CHARACTERIZATION OF A WHITE-FLOWERED WILD SOYBEAN PLANT

YIWU CHEN AND RANDALL L. NELSON\*

## Abstract

No white-flowered accession exists among the more than 1100 *Glycine soja* (Sieb. and Zucc.) accessions in the USDA Soybean Germplasm Collection, although one-third of the *Glycine max* (L.) Merr. accessions are white flowered. One white-flowered plant was found in *G. soja* accession PI 424008A growing in Stoneville, MS, in 1998. The objective of this research is to clarify the origin of this white-flowered plant. Phenotypic traits and DNA markers were used to compare the white-flowered line with the putative parental line, PI 424008A. The results from all of the data show that the white-flowered plant is very similar to PI 424008A, indicating that the white-flowered plant was produced by a mutation in PI 424008A. The white-flowered line has been added to the USDA Soybean Germplasm Collection and is designated as PI 424008C.

FLOWER COLOR IS USED in germplasm characterization, evolution studies, and cultivar identification. Within cultivated soybean (*G. max*) white and purple flowers are common. These two colors are controlled by a single gene with purple (*W1*) being dominant (Bernard and Weiss, 1973). Approximately one-third of the soybean accessions in the USDA Soybean Germplasm Collection have white flowers, whereas in the Chinese germplasm collection, the number of white and purple flowered accessions is nearly equal (Xu et al., 1999). Almost all *G. soja* accessions have purple flowers. There are a few white-flowered *G. soja* accessions in the Chinese germplasm collection (Li, 1990, 1994), but most of these white-flowered accessions have a 100-seed weight over 3.0 g, strongly indicating that these accessions are descended from recent hybridizations with *G. max*. No one has provided an explanation for why white flowers could not occur in *G. soja*, but there has been skepticism that such a type does exist.

*Glycine soja* is predominantly self-pollinating. Kiang et al. (1992) reported an average outcrossing rate of 2.3% in *G. soja* populations on the basis of the proportion of seeds with heterozygous isozyme loci. This esti-

mate is consistent with the values of *G. max* (2.4%) reported by Ahrent and Caviness (1994) and (3.0%) Chiang and Kiang (1987). However, Fujita et al. (1997) surveyed outcrossing rates for individual plants within four *G. soja* populations using isozyme markers and found mean estimates ranging from 9 to 19%. This indicates that *G. soja* may have a higher outcrossing rate than *G. max*, and this could provide a way to transfer the white flower allele from *G. max* to *G. soja*.

There is not a white-flowered accession among the more than 1100 *G. soja* accessions in the USDA Soybean Germplasm Collection. One white-flowered plant was found in PI 424008A growing in Stoneville, MS, in 1998. The objective of this research is to clarify the origin of this white-flowered plant.

## Materials and Methods

The white-flowered plant found in PI 424008A in 1998 was harvested separately. Five progenies from this plant and PI 424008A were grown in the greenhouse in spring of 1999 at Urbana, IL, for extracting DNA and increasing seeds for field evaluation in 1999. Both lines were grown in two replications for evaluation at Stoneville, MS, in 1999.

Data were collected for flowering date, maturity date, pubescence color, pubescence form, pubescence density, leaf shape index (the ratio of the maximum length to the maximum width of the leaflet), seed coat luster, and seed weight (grams/100 seeds). Seed composition measurements included protein and oil concentration, and concentrations of five fatty acids: palmitic, stearic, oleic, linoleic, and linolenic.

The LECO FP-428 Nitrogen Determinator (LECO Corp, St. Joseph, MI) was used to determine nitrogen content, and the 6.25 conversion factor was used to calculate protein concentration on a dry weight basis. A 5 MHz nuclear magnetic resonance spectrometer (Newport Oxford Instruments, Newport Pagnell, England) was used to measure oil concentration that is reported on a dry weight basis. Fatty acid methyl esters were prepared from chloroform/hexane/methanol (8:5:2, v/v/v) extracts of crushed seed by transmethylation with sodium methoxide. Fatty acid composition was measured with a Hewlett-Packard 5890-II (Palo Alto, CA) gas chromatograph equipped with dual flame ionization detectors, and a 0.53-mm by 30-m AT-Silar capillary column (Alltech Associates, Deerfield, IL). Authentic fatty acids were used for calibration.

Genomic DNA was isolated from the first trifoliate leaves of five greenhouse grown seedlings by the CTAB (hexadecyltrimethyl ammonium bromide) method of Kisha et al. (1997). Before lyophilization, harvested leaf tissue was placed in 15-mL screw-cap tubes and frozen at  $-80^{\circ}\text{C}$ . Four glass beads were added to each tube and shaken on a shaker for 3 min. The DNA concentration for polymerase chain reaction (PCR) was adjusted to 10 ng/ $\mu\text{L}$  by spectrophotometer readings at wavelengths of 260/280 nm. A total of 44 decanucleotide primers including 35 primers of a core set identified by Thompson and Nelson (1998), and nine randomly selected primers from Operon Technologies Inc. (Alameda, CA) were used in this study (Chen, 2002). The protocol of Kresovich et al. (1994) with minor modifications was used for DNA amplification. The DNA was amplified in a 9700 Applied Biosystems thermocycler (Foster City, CA) with the following amplifying regime: an initial denaturation of 2 min at  $94^{\circ}\text{C}$  followed by

Yiwu Chen, Dep. of Crop Sciences, 1101 W. Peabody Dr., University of Illinois, Urbana, IL 61801; Randall, L. Nelson, USDA-Agricultural Research Service, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Dep. of Crop Sciences, 1101 W. Peabody Dr., University of Illinois, Urbana, IL 61801. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the USDA or the University of Illinois and does not imply its approval to the exclusion of other products or vendors that may also be suitable. Received 16 Dec. 2001. \*Corresponding author (rlnelson@uiuc.edu).

Published in Crop Sci. 44:339–342 (2004).  
© Crop Science Society of America  
677 S. Segoe Rd., Madison, WI 53711 USA

**Abbreviations:** MG, Maturity Group; PI, Plant Introduction.

45 cycles of 1 min at 94°C, 5 min ramp to 72°C, and 2 min at 72°C for annealing. After the cycling was completed, the reactions were held at 72°C for 7 min before cooling to 4°C. Amplified DNA fragments were separated in 1% (w/v) agarose gels with 1× Tris-acetate buffer at 125 V constant power for 2.5 h, stained with ethidium bromide for 20 min, and visualized under UV light.

Primer pairs developed by P.B. Cregan, USDA-ARS, Beltsville, MD, were used to amplify 72 simple sequence repeats (SSR) loci that cover the 20 soybean linkage groups and because of differences in allele sizes can be multiplexed using three fluorescent dyes. Forty-eight of these primers were proposed by Song et al. (1999) as a core SSR primer set to be used in soybean cultivar identification. An additional 24 primers were added to the core SSR primer set (P.B. Cregan, personal communication, 2000). The 25-μL PCR reaction mix contained 10 mM Tris-HCl (pH 8.4); 3.0 mM MgCl<sub>2</sub>; 0.2 mM each of dATP, dCTP, dGTP, and dTTP (Sigma Chemical, St. Louis, MO); 0.5 μM primer; approximately 50 ng of genomic DNA; 1 unit of *Thermus aquaticus* (Taq) DNA polymerase (Gibco/BRL Life Technologies, Inc., Gaithersburg, MD); and sterile H<sub>2</sub>O. The DNA was amplified in a 9700 Applied Biosystems thermocycler. Cycling consisted of an initial denaturation of 4 min at 94°C followed by 33 cycles of 25 s at 94°C, 25 s annealing at 47°C, and 25 s synthesis at 68°C; and finally an additional 5 min extension at 72°C before cooling to 4°C. PCR products were diluted 1:10 in H<sub>2</sub>O and 2 μL of each diluted PCR reaction was added to 12 μL of formamide/0.5 μL of marker (GeneScan-500 ROX Size Standard). Samples of combined PCR products were separated on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, CA) using a capillary, Pop-4 Polymer, virtual filter set D, and ABI PRISM GeneScan v 3.1 software. The data were collected by ABI PRISM 310 Data Collection v 1.0.4 software.

RAPD fragments were scored as either present (1) or absent (0). To compare similarity of the two lines, Nei and Li's (1979) genetic similarity coefficient was employed to measure the genetic distance between the two genotypes by the following formula:  $S_{ij} = 2a/(2a + b + c)$ . Where  $a$  is the number of 1, 1 matches between subjects  $i$  and  $j$ ;  $b + c$  is the number of 0, 1 or 1, 0 mismatches for subject  $i$  and  $j$ . The ALLDIST procedure was applied to calculate Nei and Li's (1979) similarity coefficient in PC SAS (SAS Institute, 1999).

Using the SSR loci, we calculated a genetic dissimilarity coefficient using the formula  $1 - (\text{total score}/\text{total loci})$ , where total loci is the total number of SSR loci compared, and total score is the sum of values over those loci (Diwan and Cregan,

1997). The score for a pair of genotypes having both alleles identical at a locus is 1.0, having one identical allele is 0.5; and having no identical alleles is 0.

## Results and Discussion

PI 424008A originated from Kyonggi, South Korea, and was added to the USDA Soybean Germplasm Collection in 1976. PI 424008A is in maturity group V and has purple flowers, tawny pubescence, and black pod, seed coat, and hilum. PI 424008A and the white-flowered line are the same for all of the qualitative traits observed except flower color (Table 1). Variation for pubescence form between appressed and very appressed is not uncommon for the same genotype in different environments. The quantitative traits were similar in 1998, and statistically there were no significant differences between lines for the replicated values in 1999 (Table 1). It is highly unlikely that the white-flowered plant is the product of a hybridization with cultivated soybean because the plant has none of the characteristics intermediate between *G. soja* and *G. max* that are common in hybrid progeny. If the white-flowered plant were a hybrid between *G. soja* and *G. max*, segregation would continue in PI 424008A. In 1999, 60 plants of PI 424008A from the same source planted in 1998 were examined, and no white flowered plant was found. PI 424008A could not have crossed with a white-flowered *G. soja* plant, because no white-flowered *G. soja* existed in the USDA Soybean Germplasm Collection before the identification of this line.

Forty-four primers generated a total of 188 RAPD fragments. There were 180 fragments present in both lines, and 8 fragments were polymorphic. OPO-01 generated four polymorphic bands and OPE-01, OPH-02, OPL-09, and OPL-18 each amplified one polymorphic band. Nei and Li's similarity coefficient was 0.98. This indicates near identity in terms of genetic relationships. For comparison, there are similar data on 'Illini' and 'A.K. (Harrow)', which were selected from the same introduced seed lot, and 'Ralsoy' and 'Arksoy', which was selected from Ralsoy (Bernard et al., 1987). Both pairs of lines are nearly indistinguishable phenotypically

**Table 1. Phenotypic traits for the white-flowered *G. soja* line and the putative parent, PI 424008A.**

Characters	1998		1999†	
	Putative parent line (purple)	White-flowered line	Putative parent line (purple)	White-flowered line
Flower date	8-4	8-4	7-15	7-19
Maturity date	9-26	9-26	9-26	9-26
Pubescence color	Tawny	Tawny	Tawny	Tawny
Pubescence form	Appressed	Appressed	Very appressed	Very appressed
Pubescence density	Normal	Normal	Normal	Normal
Seed coat luster	Bloom	Bloom	Bloom	Bloom
Leaf shape (index)	2.3	2.4	2.7 a‡	2.9 a
Seed weight (g/100 seeds)	2.8	2.4	3.3 a	3.1 a
Protein (mg g <sup>-1</sup> )	468	465	429 a	402 a
Oil (mg g <sup>-1</sup> )	121	116	125 a	127 a
Palmitic acid (mg g <sup>-1</sup> )	130	126	121 a	116 a
Stearic acid (mg g <sup>-1</sup> )	38	39	38 a	39 a
Oleic acid (mg g <sup>-1</sup> )	152	154	132 a	140 a
Linoleic acid (mg g <sup>-1</sup> )	562	547	586 a	574 a
Linolenic acid (mg g <sup>-1</sup> )	118	135	125 a	131 a

† Means of two replications.

‡ Means with the same letter did not exceed the LSD value at at 0.01 probability level.

(Bernard et al., 1998; Hill et al., 2001). On the basis of data from 281 polymorphic RAPD fragments, Thompson (1997) reported that the simple matching coefficient (Sokal and Michener, 1958) of Illini-A.K. (Harrow) was 0.70 and for Ral soy-Ark soy was 0.74. The simple matching coefficient is calculated with a different formula than Nei and Li's similarity coefficient but they both estimate genetic similarity.

Sixty-seven SSR loci were homozygous and identical in both lines. Satt431 was homozygous in both lines but with different alleles in each line. Satt434 and Satt453 were heterozygous in PI 424008A, but only one allele was found at each locus in the white-flowered line. Satt510 was heterogeneous in the white-flowered line, but homozygous for one of the alleles in PI 424008A. Based on these data, the genetic similarity coefficient of SSR loci was calculated as 0.97 between the two lines. Both marker systems indicate a high degree of similarity between the two lines.

The morphological and DNA data both indicate that the white-flowered plant is nearly identical to PI 424008A. If the white-flowered plant were a mutant of PI 424008A, it would be different at only a single locus. The DNA results showed few but multiple differences. All *G. soja* accessions are purelined before they are added to the USDA Soybean Germplasm Collection. This means that each accession is derived from a single plant that came from a single seed in the original seed lot. Progenies from that plant were characterized and no obvious phenotypic segregation was observed. If any heterogeneity exists in an accession, the founder plant would have been heterozygous at that locus. It is possible that the SSR loci that were heterogeneous between the white-flowered line and PI 424008A were actually heterogeneous in PI 424008A but were not detected in the original sample of 5 plants. Twenty additional plants from PI 424008A were characterized for the four polymorphic SSR loci. The percentages of heterozygosity of Satt431, Satt434, Satt453, and Satt510 loci were 45, 35, 55, and 30%, respectively. All of the alleles found in the white-flowered plant were also present in PI 424008A. We did not confirm that differences detected between PI 424008A and the white-flowered plant with RAPD fragments could also be explained by variation within PI 424008A but that seems likely.

From these data we conclude that the white-flowered plant was produced by a mutation in PI 424008A and designate this line as PI 424008C. PI 424008C was crossed with 'Williams', a white-flowered soybean cultivar. No segregation was detected in the F<sub>2</sub> progeny indicating that the mutation was at the *W1* locus.

Skewed distributions of flower color like that in *G. soja* exist in other species, and flower color variations have been used as a model for the experimental study of evolution in the morning glory [*Ipomoea purpurea* (L.) Roth [= *Pharbitis purpurea* (L.) Voigt]] (Clegg and Durbin, 2000), in which the frequency of white-flowered plants within a wild population is significantly lower than purple flower. They found that insect pollinators discriminate against white phenotypes when white flowers are rare in populations. Since the plant is self-compatible,

pollinator bias could result in an increase in self-fertilization in white maternal plants, which should lead to an increase in the frequency of alleles conditioning white flowers according to modifier gene theory. In fact, the frequency of white flower alleles is low. Studies of geographical distribution of the gene frequency within the species demonstrated other disadvantages associated with the white phenotype. *Glycine soja* is a self-pollinated species and the process of evolution in morning glory may not be what occurred in *G. soja*, but there may be environmental or genetic factors that are influencing the allele frequency at the *W1* locus in *G. soja*.

### Acknowledgments

The authors thank Dr. Richard F. Wilson for providing the chemical composition data used in this research and Dr. Thomas C. Kilen for growing and harvesting the field plots.

### References

- Ahrent, D.K., and C.E. Caviness. 1994. Natural cross pollination of twelve soybean cultivars in Arkansas. *Crop Sci.* 34:376-378.
- Bernard, R.L., C.R. Creemeens, R.L. Cooper, F.I. Collins, O.A. Krober, K.L. Athow, F.A. Laviolette, C.J. Coble, and R.L. Nelson. 1998. Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (FC 01.547 to PI 266.807). *USDA Tech. Bull.* 1844.
- Bernard, R.L., G.A. Juvik, and R.L. Nelson. 1987. USDA soybean germplasm collection inventory. Vol. 1. International Agricultural Publications. INTSOY Series Number 30. College of Agriculture, University of Illinois, Urbana, IL.
- Bernard, R.L., and M.G. Weiss. 1973. Qualitative genetics. p. 117-154. *In* B.E. Caldwell (ed.) *Soybean: Improvement, production, and uses*. Agron. Monogr. 16. ASA, Madison, WI.
- Chen, Y. 2002. Evaluation of diversity in *Glycine soja* and genetic relationships within the subgenus *soja*. Ph.D. diss. Univ. of Illinois (Diss. Abstracts Int. Vol: 63-02, Sec. B, p. 0596), Urbana-Champaign, IL.
- Clegg, M.T., and M.L. Durbin. 2000. Flower color variation: A model for the experimental study of evolution. p. 211-234. *In* F.J. Ayala et al. (ed.) *Variation and evolution in plants and microorganisms: Toward a new synthesis 50 years after Stebbins*. National Academy Press, Washington, DC.
- Chiang, Y.C., and Y.T. Kiang. 1987. Geometric position of genotypes, honeybee foraging patterns and out-crossing in soybean. *Bot. Bull. Acad. Sinica* 28:1-11.
- Diwan, N., and P.B. Cregan. 1997. Automated sizing of fluorescent labeled simple sequence repeat markers to assay genetic variation in soybean. *Theor. Appl. Genet.* 95:723-733.
- Fujita, R., M. Ohara, K. Okazaki, and Y. Shimamoto. 1997. The extent of natural cross-pollination wild soybean (*Glycine soja*). *J. Hered.* 88:124-128.
- Hill, J.L., K. Peregrine, G.L. Sprau, C.R. Creemeens, R.L. Nelson, M.M. Kenty, T.C. Kilen, and D.A. Thomas. 2001. Evaluation of the USDA soybean germplasm collection: Maturity Groups VI to VIII (FC 03.659 to PI 567.235B). *USDA Tech. Bull.* 1894. U.S. Gov. Print. Office, Washington, DC.
- Kiang, Y.T., Y.C. Chang, and N. Kaizuma. 1992. Genetic diversity in natural populations of wild soybean in Iwate prefecture. *J. Hered.* 83:325-329.
- Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. *Crop Sci.* 37:1317-1325.
- Kresovich, S., W.F. Lamboy, R. Li, A.K. Szewc-McFadden, and S.M. Blik. 1994. Application of molecular methods and statistical analysis for discrimination of accessions and clones of vetiver grass. *Crop Sci.* 34:805-809.
- Li, F.S. 1990. Chinese *G. soja* collection catalog. China Agricultural Press, Beijing, China.

- Li, F.S. 1994. Chinese *G. soja* collection catalog (continued). China Agricultural Press, Beijing, China.
- Nei, M., and W. Li. 1979. Mathematical model for studying genetic variation in term of restriction endonucleases. *Proc. Natl. Acad. Sci. (USA)* 76:5267–5273.
- SAS Institute. 1999. SAS/STAT user's guide, Version 8.0, First ed., SAS Inst., Inc., Cary, NC.
- Sokal, R.R., and C.D. Michener. 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38:1409–1438.
- Song, Q.J., C.V. Quigley, R.L. Nelson, T.E. Carter, H.R. Boerma, J.L. Strachan, and P.B. Cregan. 1999. A selected set of trinucleotide simple sequence repeat markers for soybean cultivar identification. *Plant Var. Seed.* 12:207–220.
- Thompson, J. A. 1997. Identification and utilization of diverse germplasm for soybean yield improvement. Ph.D. diss. (Diss Abstr Inter Vol: 58–06, Section: B, page: 2765). University of Illinois, Urbana-Champaign.
- Thompson, J.A., and R.L. Nelson. 1998. Core set of primers to evaluate genetic diversity in soybean. *Crop Sci.* 38:1356–1362.
- Xu, Z.Y., R.Z. Chang, L.J. Qiu, J.Y. Sun, and X.H. Li. 1999. Evaluation of soybean germplasm in China. p.156–165. *Proceedings of Global Soy Forum 99. World Soybean Research Conference, VI, Chicago, IL. 4–7 Aug. 1999.*